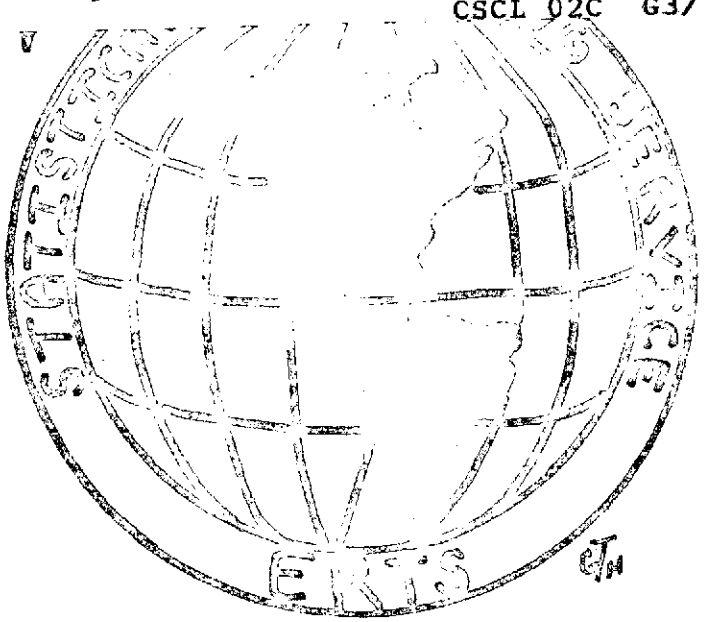
Type I Progress Report

February 20, 1974 - April 19, 1974

Crop Identification & Acreage Measurement Utilizing ERTS Imagery 013



Principal Investigator

Donald H. Von Steen AG 328

Table of Contents

Analysis of Aerial Photography

Scanning of the RC-8 high altitude photography from mission 208, flown by NASA on August 18, 1972, has been completed. This photography covers flightlines 3 and 10 in Kansas. It was found that locating fields within segments and recording their coordinates were time-consuming tasks. Sketches of each segment were drawn showing each field within the segment. Corner coordinates of rectangular fields were then recorded on the sketch from the DCRS (digital coordinate readout system) on the microdensitometer system. For irregular shaped fields (non-rectangular), as many as 10 boundary coordinates were recorded. Figure 1 is a simplified example of a sketch of a segment with field boundary coordinates recorded. The segments were then scanned by the microdensitometer with an effective aperture of 240 microns square, and the optical densities and percent transmission values were recorded on magnetic tapes for each of the four color filters (red, green, blue, and clear).

Conversion of the microdensitometer scanning data into a SAS compatable format is currently underway using the PDSCMS computer program. A computer program has been developed to compute the location of and extract the data for each field within a segment. The program operates in conjunction with SAS to extract rectangular fields parallel to the microdensitometer scanning axes. Irregular shaped fields are subdivided into several rectangular fields parallel to the scanning axes. Recording the keypunching input data for the field extraction program is currently underway.

Cost Analysis

The following is a breakdown of approximate time and cost involved in scanning the aircraft photography in Kansas, and converting the data into a format suitable for crop classification.

Activity	Average Time/Segment
Sketch segment and record field boundaries	37 min.
Microdensitometer Scanning	33 min.
Recording and Keypunching input data for field extraction	40 min.

Total man hours	=	1.83 hours
Approx. cost/man hour	=	$4.50
Average cost/segment	=	1.83 x 4.50 = $8.23

ADP costs on a per segment basis are as follows:

PDSCMS data conversion		$12.00
Field extraction		$15.00
	Total	$27.00

Thus, the average cost per segment for scanning the aircraft photography and converting the data into a format suitable for crop classification is $35.23.

Simplified Sketch of Segment with Field Boundary Coordinates Recorded

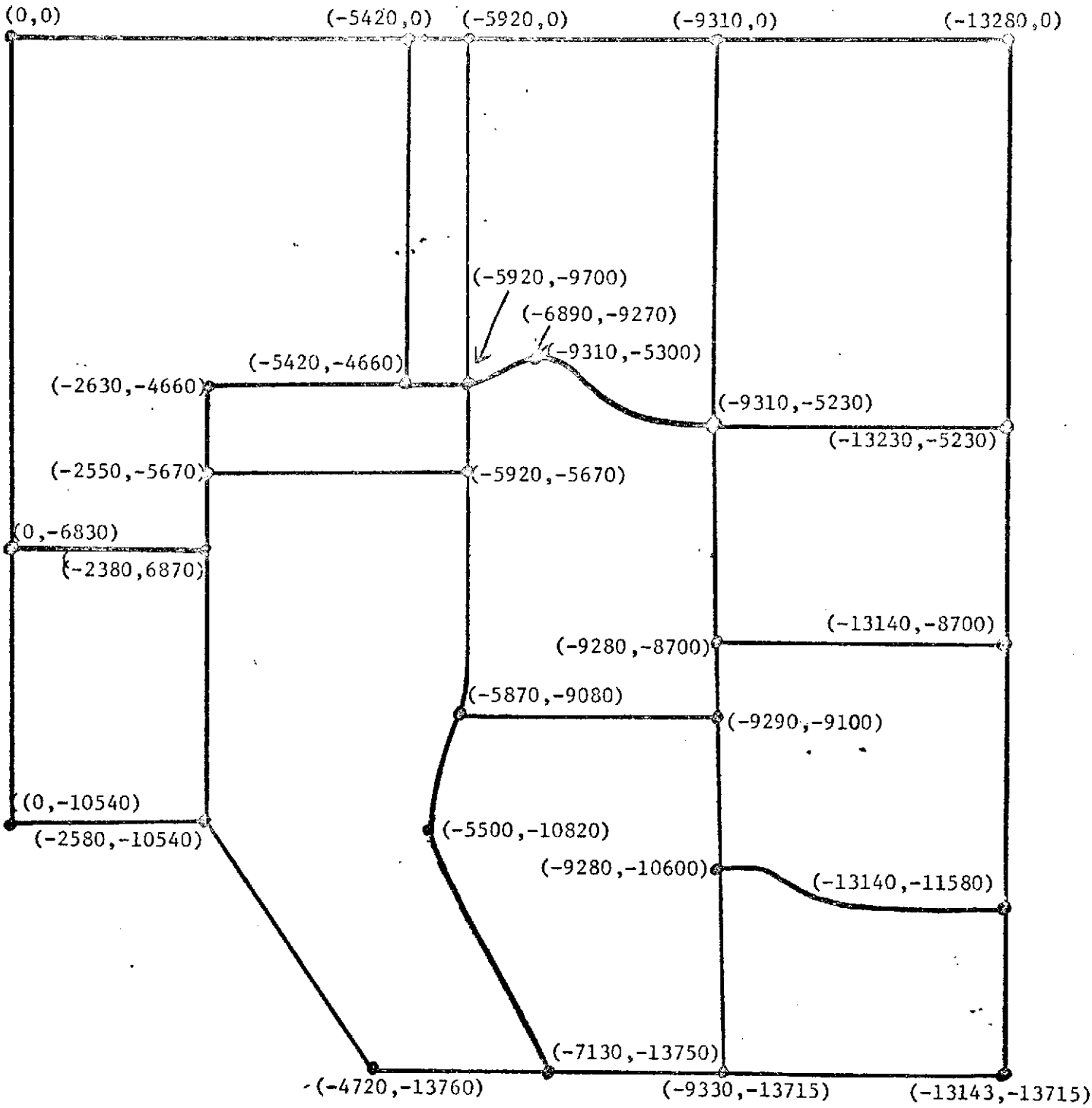


Figure 1.

Analysis of ERTS Data in Idaho

Classification results in Idaho were poor. One reason for the poor results seemed to be a banding problem. Upon our request, NASA reprocessed Idaho frame 1035-17525 to remove the banding. Classification was done using the reprocessed tapes and identical results were obtained as were reported previously.

Analysis of ERTS Data in Kansas Test Site

Objectives

The objectives of analysis of ERTS data in Kansas are:

A. computation of classification rates for the Kansas test site.

B. computation of correlation coefficients between ground truth acreage and classified pixels.

C. study the effects of classification in one ERTS frame using training parameters from an adjacent pass.

D. study the classification of a Kansas county.

Approach

A. ERTS imagery, for the area of interest, was too cloudy to be useful prior to September 21, 1972. The study was made on September 21 and 22, imagery. The area of interest in Kansas was divided by two ERTS passes thus the training data was also divided. Twenty-two segments were in the September 21, imagery. Seven of these segments were hidden by clouds. Therefore, 15 segments were used as training and test data.

Since the time of year was not optimum, a visual inspection of the gray scale printout of MSS band 5 and ground truth were used to select particular fields to use as training fields i.e. those fields which were partially harvested and those with a confusion of symbols were discarded. Another reason for selecting fields was to compare parameters from one pass with those from another as described later in this report. The "select fields" were used for both training and classifying. The classification based on these select fields is presented in Table 1. The overall performance was 91.2%. The classification used the standard pointwise quadratic discriminant functions found in LARSYS with the added feature of unequal prior probabilities. The unequal prior probabilities use information that is available about the likelihood of certain crops. If, for example, corn is more likely to be encountered than grain sorghum, corn is given a high chance of occurrence. In most classification using unequal prior probabilities done in Kansas, the weights were:

1.	Alfalfa	.03
2.	Pasture	.72
3.	Corn	.09
4.	Grain Sorghum	.16

Prior probabilities in this report were computed from data gathered by the Statistical Reporting Service in early June 1972 (June Enumerative Survey).

Table 1--Classification matrix for September 21, 1972 imagery (MSS bands 4,5,6, 7) using quadratic discriminant functions with unequal prior probabilities in Kansas test site for select fields.

Class	No. of sample points	Percent Correct	Number of samples classified into				
			Alfalfa	Pasture	Corn	Grain Sorghum	Threshold
Alfalfa.....	63	100.0	63	0	0	0	0
Pasture.....	172	98.3	0	169	2	1	0
Corn.....	51	90.2	0	1	46	4	0
Grain Sorghum..	78	69.2	0	10	14	54	0
Total.....	364		63	180	62	59	0

Overall performance = 91.2

In Table 1, the number of pixels to be classified are not proportional to the prior probabilities selected. The prior probabilities are based on acreage of all segments in the Crop Reporting District, and not the segments in frame 1060-16512. Development of proper weights for areas divided by ERTS passes presents additional problems. A better correspondence would have resulted in higher overall classification, however, 91.2% is very good.

A classification was then done using all identifiable fields in the 15 segments. The results of this classification are presented in Table 2. The overall performance was 90.2%.

There was a small decrease in overall performance between Table 1 and Table 2. However, a random sample of ground truth yields a better representation and allows more statistical procedures to be applied.

The second pass required to cover the Kansas test site was analyzed in the same way as described above. The second scene contained 23 segments, but one of these segments fell in a non-agricultural area. In addition to the random segments, two additional segments were selected which contained sugar beets.

Table 3 presents the classification of "select fields" for the second pass. The fields were selected from the gray scale printout as described above. The overall performance was 75.5%.

Table 4 represents a classification of the second scene using all identifiable fields. The overall performance was 65.8%. This decrease in performance could be attributed to several things. The number of crops being classified was increased from four to seven. Increasing the number of crops will reduce the performance. Secondly, there was a confusion between most crops and pasture. This could have resulted from using late September imagery and the weight given to pasture.

Table 5 is a classification study using the same select training fields as were used in Table 3. However, in Table 5 equal prior probabilities were applied. In Table 5, the overall performance at 79.2% is actually better than the 75.5% in Table 3. Applying weights based on all fields to a non-random selection of fields in a particular area is the cause for the lower classification in Table 3.

Table 6 presents an unweighted classification of all identifiable fields in scene 1061-16570. This table is comparable with the weighted classification presented in Table 4. The overall performance was increased 4.4% by using prior probabilities. When all fields are used in the classification, the total acres per crop more closely estimate the true prior probabilities of the model.

Table 2--Classification matrix for September 21, 1972 imagery (MSS bands 4,5,6, 7) using quadratic discriminant functions with unequal prior probabilities in Kansas test site.

Class	No. of sample points	Percent Correct	Number of samples classified into				
			Alfalfa	Pasture	Corn	Grain Sorghum	Threshold
Alfalfa.....	43	93.0	40	2	0	1	0
Pasture.....	6261	95.0	23	5949	121	139	29
Corn.....	332	37.7	38	110	125	59	0
Grain Sorghum..	508	64.8	38	77	60	329	4
Total.....	7144		139	6138	306	528	33

Overall performance 90.2

Table 3--Classification matrix for September 22, 1972 imagery (MSS bands 4,5,6, 7) using quadratic discriminant functions with unequal prior probabilities in Kansas test site for select fields.

Class	No. of sample points	Percent correct	Number of samples classified into				
			Alfalfa	Pasture	Corn	Grain Sorghum	Threshold
Alfalfa.....	78	84.6	66	12	0	0	0
Pasture.....	230	93.0	0	214	11	5	0
Corn.....	337	65.0	0	93	219	25	0
Grain Sorghum...	177	68.9	3	34	18	122	0
Total.....	822		69	353	248	152	0

Overall performance = 75.5

Table 4--Classification matrix for September 22, 1972 imagery (MSS bands 4,5,6,7) using quadratic discriminant functions with unequal prior probabilities in Kansas test site.

Class	No. of sample points	Percent Correct	Number of samples classified into							
			Alfalfa	Pasture	Corn	Grain Sorghum	Winter Wheat	Fallow	Sugar Beet	Threshold
Alfalfa	287	56.4	162	57	12	23	16	11	6	0
Pasture	4975	90.6	19	4508	45	44	156	180	0	23
Corn	1698	40.8	1	684	693	174	99	47	0	0
Grain Sorghum	2869	55.3	89	300	357	1586	265	268	0	4
Winter Wheat	863	13.3	14	431	16	41	115	242	0	4
Fallow	1508	64.6	10	285	44	56	134	974	2	3
Sugar Beet	25	0.0	16	2	1	1	5	0	0	0
Total	12225		311	6267	1168	1925	790	1722	8	34

Overall performance = 65.8

Table 5--Classification matrix for September 22, 1972 imagery (MSS bands 4,5,6,7) using quadratic discriminant functions with equal prior probabilities in Kansas test site for select fields.

Class	No. of sample points	Percent Correct	Number of samples classified into				
			Alfalfa	Pasture	Corn	Grain Sorghum	Threshold
Alfalfa.....	78	84.6	66	11	0	1	0
Pasture.....	230	75.2	3	173	38	16	0
Corn.....	337	87.5	0	29	295	13	0
Grain Sorghum...	117	66.1	14	16	30	117	0
Totals.....	822		83	229	363	147	0

Overall performance = 79.2

Table 6--Classification matrix for September 22, 1972 imagery (MSS bands 4,5,6,7) using quadratic discriminant functions with equal prior probabilities in Kansas test site.

Class	No. of sample points	Percent Correct	Number of samples classified into							
			Alfalfa	Pasture	Corn	Grain Sorghum	Winter Wheat	Fallow	Sugar Beet	Threshold
Alfalfa	287	50.5	145	18	30	9	24	4	57	0
Pasture	4975	80.1	61	3986	371	66	340	106	22	23
Corn	1698	70.3	80	267	1193	69	39	32	18	0
Grain Sorghum	2869	42.1	496	115	620	1209	149	103	174	3
Winter Wheat	863	23.4	20	350	50	44	202	149	44	4
Fallow	1508	50.5	18	208	79	120	256	762	62	3
Sugar Beet	25	56.0	6	2	2	0	1	0	14	0
Total	12225		826	4946	2345	1517	1011	1156	391	33

Overall performance = 61.4

The increase caused by using unequal prior probabilities in Kansas was not as great as it had been in other areas. The smaller gain from prior probabilities is perhaps caused by the fact that the ERTS data contained more information i.e. the classes were more separable. Thus, the expected gain from prior probabilities is more in areas where classification is poorer.

B. The correlation between acres and pixels were calculated. Coordinates of ground truth segments were carefully defined. The training data from each scene were used to classify the segments in that scene. The classified pixels in the two scenes were then combined (i.e. Tables 2 and 4 were combined) and correlations with known ground truth acreage were computed.

Correlations between acreage and pixels were calculated as follows:

Total acreage vs Total Pixel	$r^2 = .88$	$r = .94$
Pasture acreage vs Pasture Pixel	$r^2 = .84$	$r = .92$
Corn acreage vs Corn Pixel	$r^2 = .62$	$r = .79$
Grain Sorghum vs Grain Sorghum Pixel	$r^2 = .58$	$r = .76$

When pixels and acreage are this highly correlated, remotely sensed data is beneficial.

C. In this study, the statistics compiled on one ERTS frame were used to classify points in the adjacent frame. As described earlier, two adjacent ERTS passes were used to obtain necessary coverage of Kansas. The "select fields" from both scenes (as described in Section A), had four classes (alfalfa, pasture, corn, grain sorghum). These four classes were also the classes for the "all fields" in frame 1060-16512. One requirement is that the same classes be used for training as those classified. The classification used the quadratic discriminant function with unequal prior probabilities.

Table 7 presents the results of classifying the "select fields" in ERTS frame 1060-16512 using training statistics generated from "select fields" in frames 1061-16570. The overall performance was 54.4%. however, the average performance by classes1/ was 33.3% correct classification. The 100% correct classification of the pasture class greatly influenced the overall classification.

1/ The average performance by classes is computed by averaging the percent correctly identified for each class.

Table 7--Classification matrix of "select fields" in frame 1060-16512 classified using statistics generated from "select fields" in frame 1061-16570.

Class	No. of sample points	Percent Correct	Number of samples classified into				
			Alfalfa	Pasture	Corn	Grain Sorghum	Threshold
Alfalfa....	63	0.0	0	61	0	1	1
Pasture....	172	100.0	0	172	0	0	0
Corn.....	51	0.0	3	7	0	41	0
Grain Sorghum....	78	33.3	7	28	15	26	2
Total.....	364		10	268	15	68	3

Overall performance = 54.4

Table 8 is a classification of all identifiable fields in the segments in frame 1060-16512, using the statistics generated from the "select fields" in frame 1061-16570. The classification with an overall performance of 85.5% and an average class performance of 43.5% is very good. Here again, it was the correctly classified pasture points which kept the averages high. In Table 8, more fields were classified and the influence of prior probabilities was more beneficial than in the cases where select fields were classified.

Table 9 shows a classification of "select fields" in frames 1061-16570 using statistics generated from "all fields" in frame 1060-16512. In this study, the overall performance slipped to 49.0% but the average class performance was 59.1%. Classification was very good in all classes except corn which was confused with pasture and grain sorghum. The time of year may have caused this confusion.

D. The border of Stevens County, Kansas was drawn on a gray scale map of MSS band 5. The area was then defined on punch cards and classified. Training data for the classification was obtained from segments in the Crop Reporting District which contains Stevens County. Three of these segments were actually in Stevens County. A total of 410,505 pixels were classified which correspond to a calculated 466,560 acres in the county.

Alfalfa, pasture, corn, and grain sorghum were the crops classified. The following classification was obtained:

Number of Pixels	Alfalfa	Pasture	Corn	Grain Sorghum	Threshold
410,505	5,362	172,021	30,448	165,107	37,567
	1.3%	41.9%	7.4%	40.2%	9.2%

The prior probabilities as a percentage which were applied were the following:

Alfalfa	3%
Pasture	72%
Corn	9%
Grain Sorghum	16%

There is an indication of confusion between pasture and grain sorghum. Ways to use this data to produce a final estimate are being investigated.

Table 8--Classification matrix of "all fields" in frame 1060-16512 classified using statistics generated from "select fields" in frame 1061-16570.

Class	No. of sample points	Percent Correct	Number of samples classified into				
			Alfalfa	Pasture	Corn	Grain Sorghum	Threshold
Alfalfa....	43	65.1	28	3	0	12	0
Pasture....	7229	93.2	8	6736	12	314	159
Corn.....	332	7.5	8	79	25	204	16
Grain Sorghum....	508	28.3	16	105	75	144	168
Total.....	8112		60	6923	112	674	343

Overall performance = 85.5

Table 9--Classification matrix of "select fields" in frame 1061-16570 classified using statistics generated from "all fields" in frame 1060-16512.

Class	No. of sample points	Percent Correct	Number of samples classified into				
			Alfalfa	Pasture	Corn	Grain Sorghum	Threshold
Alfalfa....	78	80.8	63	12	0	0	3
Pasture....	230	94.3	0	217	4	8	1
Corn.....	337	9.2	5	140	31	161	0
Grain Sorghum....	177	52.0	12	30	43	92	0
Total.....	822		80	399	78	261	4

Overall performance = 49.0

Summary

A. Classification in Kansas was good even though the time of year was not optimum. The use of prior probabilities increased the overall performance of the classifier. However, the increase was not as great as in areas where classification was poorer. In Kansas, selection of fields visually from a gray scale printout of MSS band 5 and ground truth increased classification, but reduced the effectiveness of some statistical procedures.

B. There was a good correlation between pixels and acreage in Kansas. The correlations ranged from .76 to .94, satellite data would be beneficial when used as supplemental information in a regression estimator and could reduce the sampling error by 58% to 80%.

C. Classification in one ERTS frame using statistics generated in an adjacent frame were better than expected. The exercise was conducted as an experiment, and the authors do not imply signature extension could be applied in general. It has not been shown that in general one would expect good or bad results from such a practice. However, the results obtained here do show hope in the area of using training data in one frame to classify in others. The ability to use training data in more than one frame would be of great benefit when working with a very large area. Calibration changes between ERTS frames do create problems that must be dealt with.

D. Stevens County, Kansas was classified and the total number of pixels in each crop category was counted. An investigation of various ways to produce estimates from these counts is underway.

Analysis of ERTS Data in South Dakota

Objectives

The objective of this study was to determine the classification rate in the South Dakota test site.

Approach

Imagery for three dates was available. However, the August and early September imagery was too cloudy to be useful. Thus, late September imagery was used. All 34 segments were contained in one ERTS frame (1060-16491). The segments and fields within segments were located and defined on punch cards. These segments were used for both training and classifying.

The LARS classifier was used with the addition of prior probabilities to the model. The classifier is a standard parametric discriminant analysis routine.

Table 10 presents a weighted classification of the fields in all segments in South Dakota. The overall performance was 30%, but the average class performance was 15%. Almost all classes in Table 10 were classified as either pasture or oats. This would indicate that in late September all classes look very much alike. Plot 1 shows a Coincident Spectral Plot for the ten classes. It is clear that there is almost no separation of the classes in any of the 4 MSS bands. It would be impossible to separate these classes with these data. One should remember in examining Plot 1 that in the true multivariate sense the ability to separate groups may not be as poor as it would appear.

Plots 2-11 show the Spectral Plot for each of the individual 10 classes. All classes look very much alike and there is very little information for the discriminant analysis.

The prior probabilities used in the model gave most of the weights to the two classes oats and pasture. There was very little information for separating classes, so the results obtained were highly dependent on the prior probabilities.

Table 10--Classification matrix for September 21, 1972 imagery (MSS bands 4,5,6,7) using quadratic discriminant functions with unequal prior probabilities in South Dakota test site.

Class	No. of sample points	Percent Correct	Number of samples classified into										
			Corn	Pasture	Oats	Barley	Rye	Alfalfa	Flax	Sudex	Idle	Fallow	Threshold
Corn	1060	0.1	1	753	275	3	0	0	3	0	12	10	3
Pasture	812	88.4	1	718	86	1	0	0	0	0	2	4	0
Oats	243	40.3	0	142	98	0	0	0	0	0	0	3	0
Barley	97	0.0	0	77	17	0	0	0	1	0	2	0	0
Rye	16	0.0	0	15	1	0	0	0	0	0	0	0	0
Alfalfa	303	0.3	0	243	51	0	1	1	0	0	0	6	1
Flax	71	4.2	0	45	23	0	0	0	3	0	0	0	0
Sudex	55	0.0	0	47	7	0	0	0	0	0	0	1	0
Idle	18	10.5	0	14	3	0	0	0	0	0	2	0	0
Fallow	82	4.9	0	59	17	0	0	0	0	0	2	4	0
Total	2758		2	2113	578	4	1	1	7	0	20	28	4

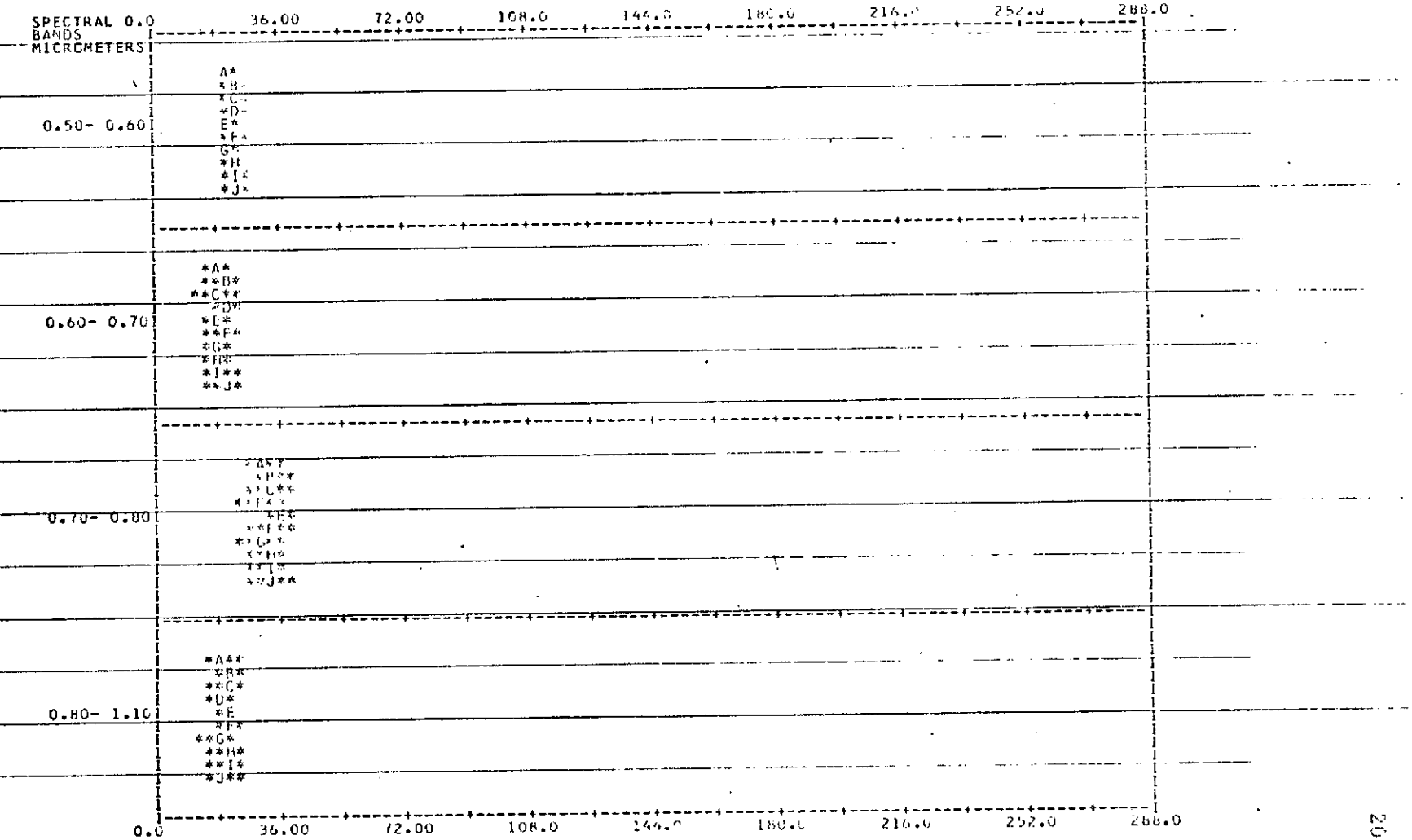
Overall performance = 30.0

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

MAR 7,1774
11 55 43 PM
LARSYS VERSION 3

COINCIDENT SPECTRAL PLOT (MEAN PLUS AND MINUS ONE STD. DEV.) FOR CLASS(ES)

LEGEND
A = CLASS 1 CORN
B = CLASS 2 OATS
C = CLASS 3 OATS
D = CLASS 4 BARLEY
E = CLASS 5 RYE
F = CLASS 6 ALFA
G = CLASS 7 FLAX
H = CLASS 8 SUDX
I = CLASS 9 IDLE
J = CLASS 10 WFALL



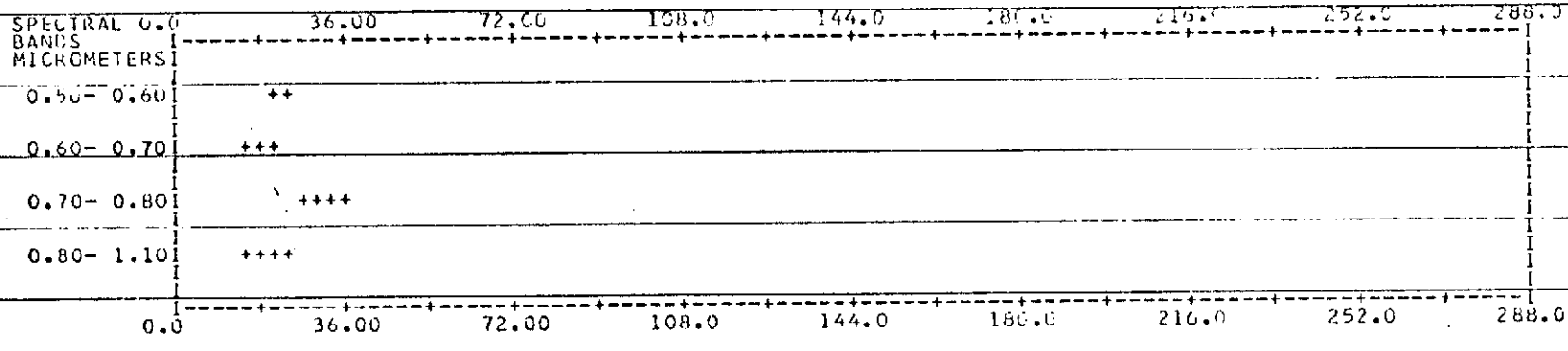
PLOT #1

An attempt to improve the classification results was made by selecting fields in the same way as described in the Kansas analysis. These selected fields were used as training data and then classified. The results of this classification are presented in Table 11. The overall performance was 26% and the average class performance was 44%. Plot 12 is the Coincident Spectral Plot for the five classes. There is very little information in the data which would aid separation of classes. The influence of the prior probabilities again was the reason pasture and oats had high correct classification rates.

CLASS....CORN

TOTAL NUMBER OF SAMPLES... 1060

SPECTRAL PLOT (MEAN PLUS AND MINUS ONE STD. DEV.)



PLOT #2

SRS
WIGTON

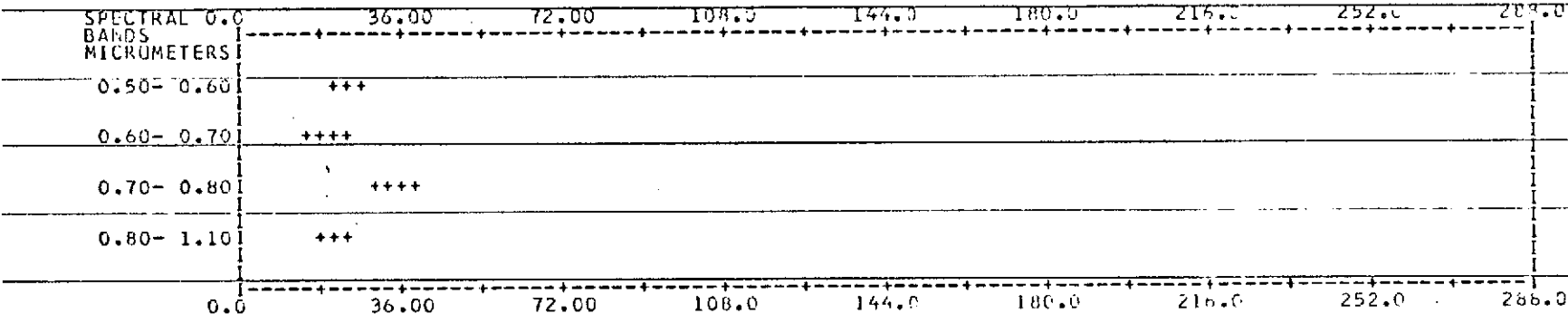
LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

MAR 7,1774
10 38 11 PM
LARSYS VERSION 3

CLASS....PAST

TOTAL NUMBER OF SAMPLES... 812

SPECTRAL PLOT (MEAN PLUS AND MINUS ONE STD. DEV.)



PLOT #3

SRS
WIGTON

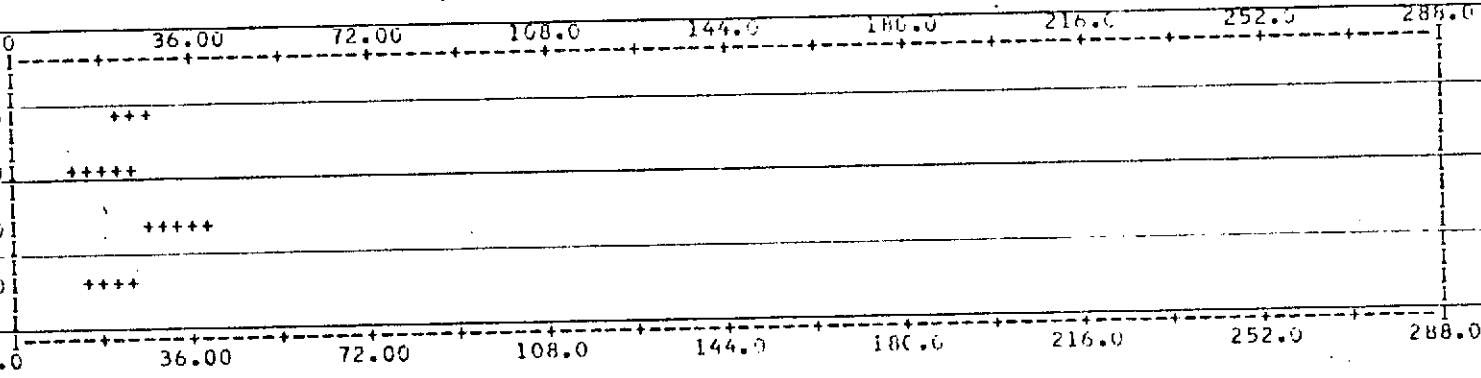
LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

MAR 7,1974
11 11 37 PM
LARSYS VERSION 3

CLASS....OATS

TOTAL NUMBER OF SAMPLES... 243

SPECTRAL PLOT (MEAN PLUS AND MINUS ONE STD. DEV.)

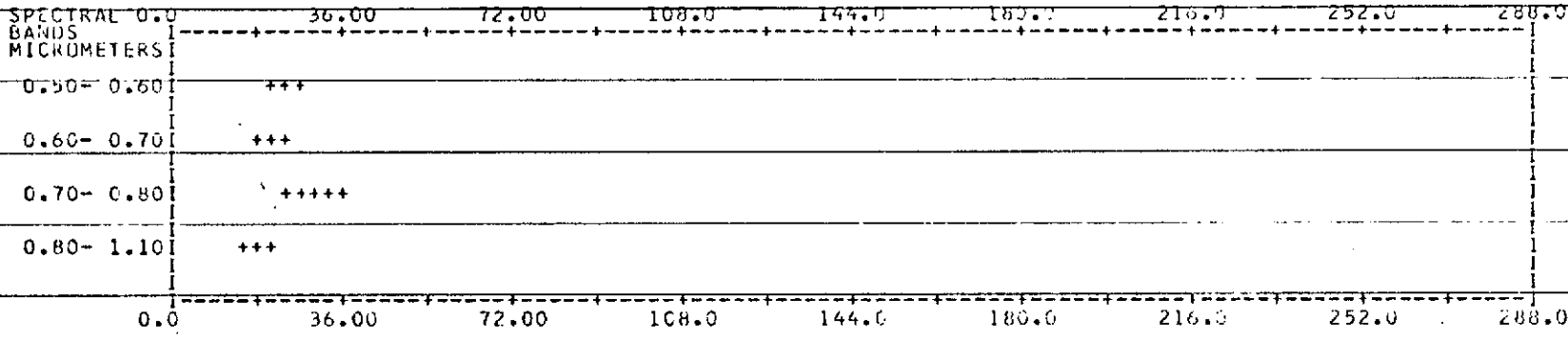


PLOT # 4

CLASS....BARLEY

TOTAL NUMBER OF SAMPLES... 97

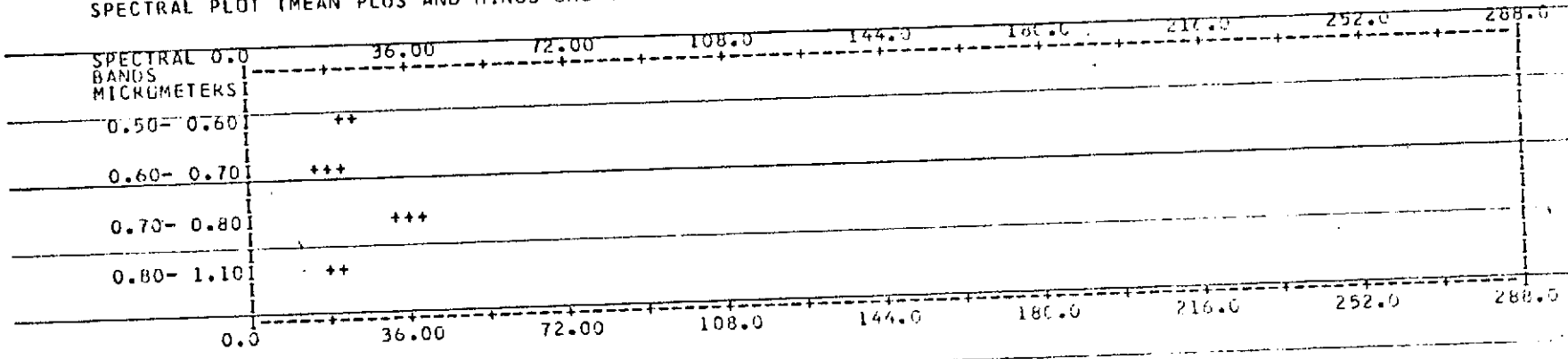
SPECTRAL PLOT (MEAN PLUS AND MINUS ONE STD. DEV.)



PLOT #5

CLASS....RYE TOTAL NUMBER OF SAMPLES... 16

SPECTRAL PLOT (MEAN PLUS AND MINUS ONE STD. DEV.)



PLOT #6

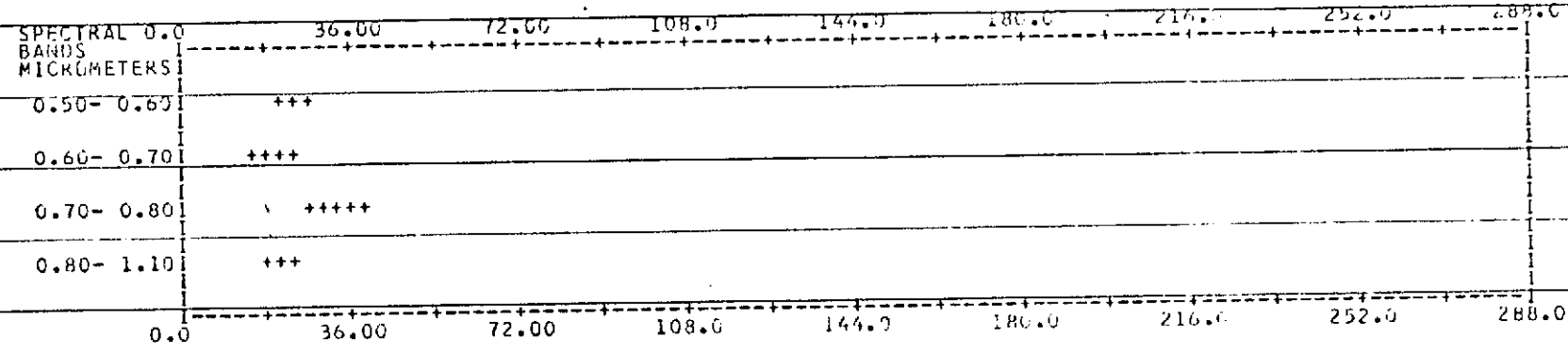
LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

MAR 7,1774
11 43 47 PM
LARSYS VERSION 3

CLASS....ALFA

TOTAL NUMBER OF SAMPLES... 303

SPECTRAL PLOT (MEAN PLUS AND MINUS ONE STD. DEV.)

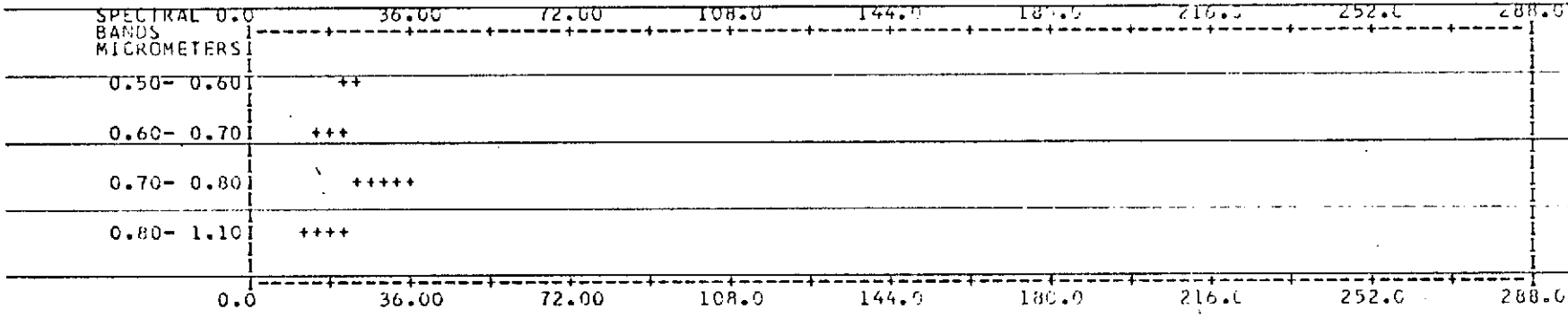


PLOT #7

CLASS....FLAX

TOTAL NUMBER OF SAMPLES... 71

SPECTRAL PLOT (MEAN PLUS AND MINUS ONE STD. DEV.)

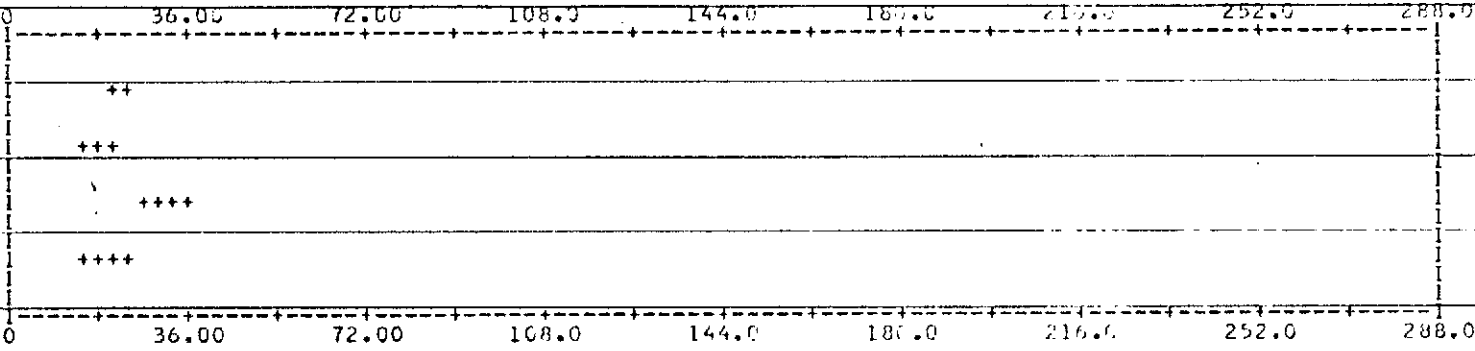


PLOT #8

CLASS....SUDX

TOTAL NUMBER OF SAMPLES... 55

SPECTRAL PLOT (MEAN PLUS AND MINUS ONE STD. DEV.)

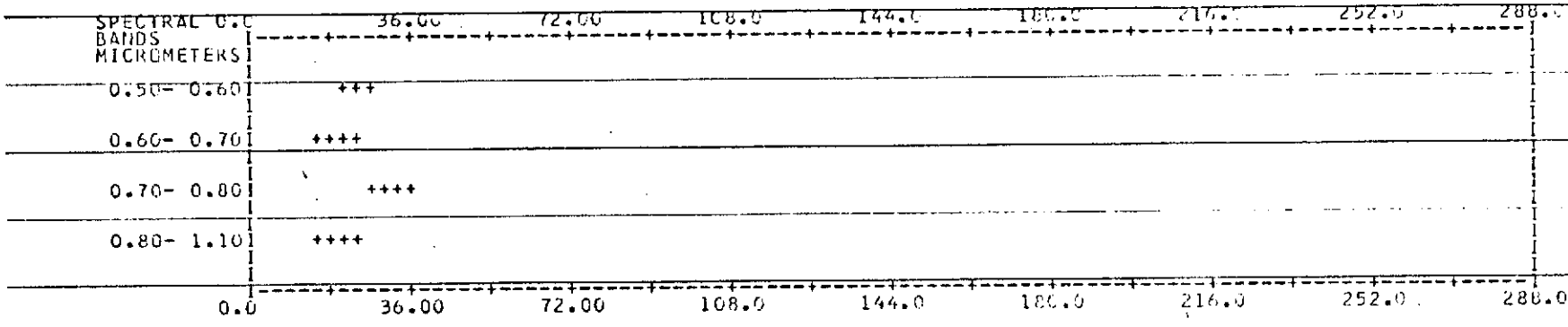


PLOT #9

CLASS....IDLE

TOTAL NUMBER OF SAMPLES... 19

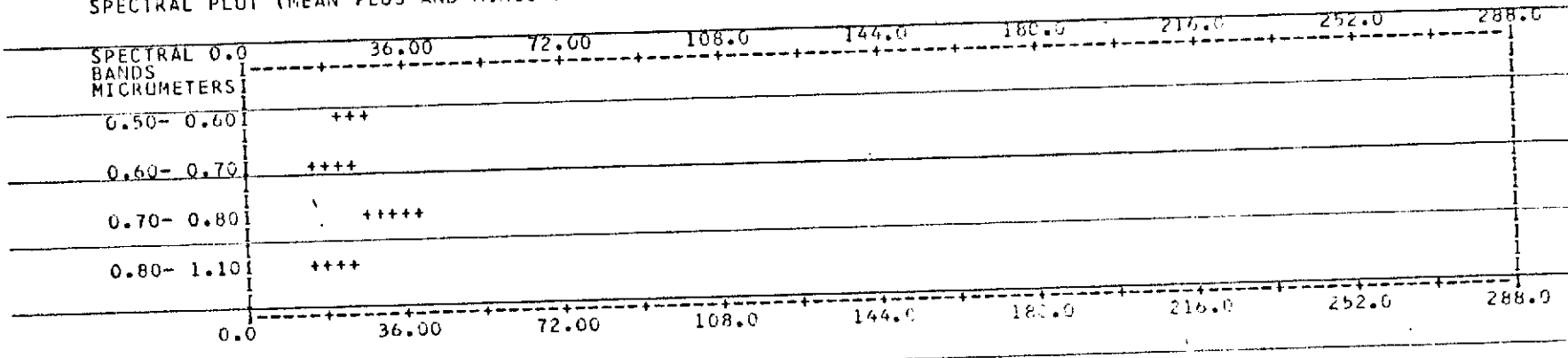
SPECTRAL PLOT (MEAN PLUS AND MINUS ONE STD. DEV.)



PLOT # 10

CLASS....WFALL TOTAL NUMBER OF SAMPLES... 82

SPECTRAL PLOT (MEAN PLUS AND MINUS ONE STD. DEV.)



PLOT #11

Table 11--Classification matrix for September 12, 1972 imagery (MSS bands 4,5,6, 7) using quadratic discriminant functions with unequal prior probabilities in South Dakota test site for select fields.

Class	No. of sample points	Percent Correct	Number of samples classified into					
			Corn	Pasture	Oats	Alfalfa	Sudex	Threshold
Corn.....	237	6.8	16	150	54	17	0	0
Pasture....	75	88.0	0	66	7	2	0	0
Oats.....	12	100.0	0	0	12	0	0	0
Alfalfa....	110	25.5	1	56	24	28	0	1
Sudex.....	36	0.0	0	30	6	0	0	0
Total.....	470		17	302	103	47	0	1

Overall performance = 26.0

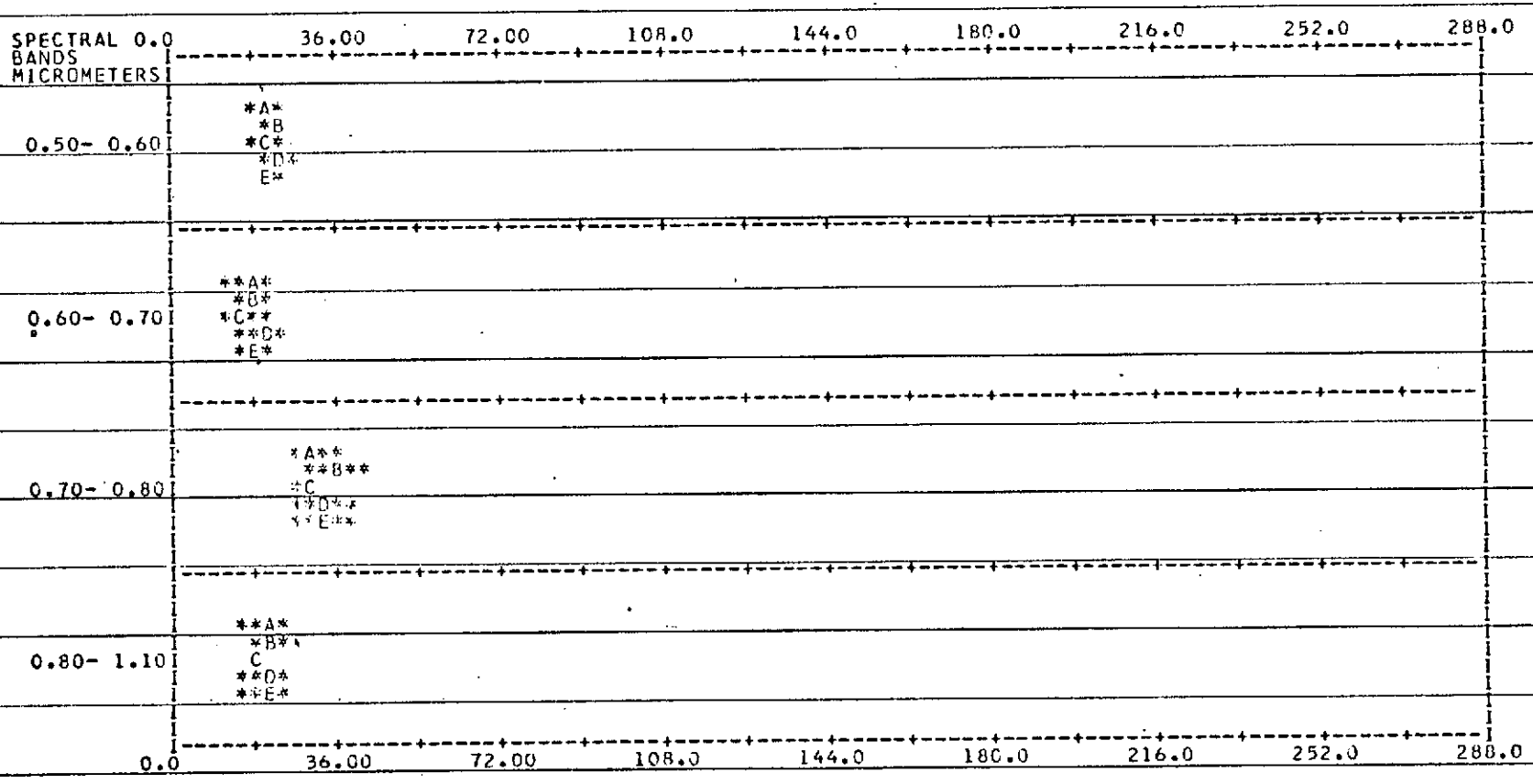
SRS
WIGTON

LABORATORY FOR APPLICATIONS OF REMOTE SENSING
PURDUE UNIVERSITY

MAR 7,1974
4 11 51 AM
LARSYS VERSION 3

COINCIDENT SPECTRAL PLOT (MEAN PLUS AND MINUS ONE STD. DEV.) FOR CLASS(ES)

LEGEND
A = CLASS 1 CORN
B = CLASS 2 PAST
C = CLASS 3 OATS
D = CLASS 4 ALFA
E = CLASS 5 SUDX



PLOT #12

Summary

In South Dakota, late September imagery was used because of cloud cover on earlier imagery. Classification results were poor. Examination of the Coincident Spectral Plot showed very little information in the ERTS data for the separation of the classes of interest. This late in the season, crops were classified as either oats or pasture.

The use of fields selected from gray scale printouts and ground truth did not improve classification, but actually reduced the overall performance.